

Consensus of Dependent Opinions

Sujoy Chatterjee¹, Anirban Mukhopadhyay¹ and Malay Bhattacharyya²

¹Department of CSE, University of Kalyani, Nadia – 741235, India

E-mail: {sujoy, anirban}@klyuniv.ac.in

²Department of IT, IEST, Shibpur, Howrah – 711103, India

E-mail: malaybhattacharyya@it.iests.ac.in

Abstract

Providing opinions through labeling of images, tweets, etc. have drawn immense interest in crowdsourcing markets. This invokes a major challenge of aggregating multiple opinions received from different crowd workers for deriving the final judgment. Generally, opinion aggregation models deal with independent opinions, which are given unanimously and are not visible to all. However, in many real-life cases, it is required to make the opinions public as soon as they are received. This makes the opinions dependent and might incorporate some bias. In this paper, we address a novel problem, hereafter denoted as dependent judgment analysis, and discuss the requirements for developing an appropriate model to deal with this problem. The challenge remains to be improving the consensus by revealing true opinions.

Introduction

Opining about a large set of questions (may be images, tweets, etc.) within a bounded time is a challenging job. It highly demands for a distributed processing. Recently, it has been understood that such problems can be easily solved with human intervention and crowdsourcing is one such successful model (Bhattacharyya 2013) (Chatterjee and Bhattacharyya 2015), (Kajino, Tsuboi, and Kashima 2012), (Jung and Lease 2012). Still, there remain a few challenges like assigning the appropriate domain experts as crowd workers, assuring the quality of opinions, etc. Due to the lack expertise or dishonest interest of crowd workers, the opinions have a high rate of noise (Raykar and Yu 2011). To obtain a better consensus of the given opinions, aggregation models are frequently used (Sheshadri and Lease 2013). In general, opinion aggregation models deal with independent opinions. These are given unanimously and are not visible to everyone. However, in many real-life scenarios, it is required to make the opinions public. The crowd workers should be able to choose the questions (in which they feel they have sufficient knowledge) to be answered. These are possible with an open crowdsourcing platform that makes the collection of opinions faster and transparent (Bhattacharyya 2014). In contrast to investigating the impact

of review bias as was studied earlier (Krishnan et al. 2014), we aim here to derive judgments from dependent opinions. The limited existing approaches in this direction based on incentive mechanisms for unknown objective truth (Prelec 2004) and debiasing methods for consensus tasks (Kamar, Kapoor, and Horvitz 2015) are not appropriate for this open model problem.

If the opinions are made open to all the crowd workers, there is a high chance of inclusion of bias in the opinions. As the opinions become dependent, aggregating them introduces a new type of judgment analysis problem. Finding the consensus of dependent crowd opinions leads to a relatively new research direction of judgment analysis and there is hardly any study available in the literature. In this paper, we address this new research direction of dependent crowd judgment analysis problem. Principally, in this problem the opinions arrive with a time interval and become visible to all the crowd workers. So, the opinions provided are an ordered (dependent) set of opinions. As the crowd workers may notice others' opinions, therefore, some preconceived notions might influence their opinions.

If a crowd worker realizes that his true opinion (the opinion in mind) is wide dissimilar from the majority, he may change his opinion (the opinion disclosed) towards the majority. A crowd worker sticking to his actual opinion, even after observing others' opinions, establishes that he is confident about his opinion. On the contrary, if a crowd worker deviates from his true opinion by a high scale, it reflects that he is not much confident. A crowd worker who is less confident about his opinions might not be efficient. Therefore, how far a crowd worker differs from others should be taken into account. From this perspective, we discuss about some novel terms. These include drop of confidence (absolute difference of a crowd worker's actual and disclosed opinion), reliability, and accuracy (how dissimilar the crowd worker's opinion is from the rest of the workers) of the crowd worker's opinion. Ultimately, these can be used as weights to derive the aggregated judgment of a question.

Motivation

The problem of finding consensus of dependent opinions is basically motivated from a recently proposed crowd-powered reviewing model. This model collects the review score for research papers (submitted in video format) from

* Works-in-Progress, Fourth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2016), Austin, TX, USA. Copyright © 2016, Retained by the authors. All rights reserved.

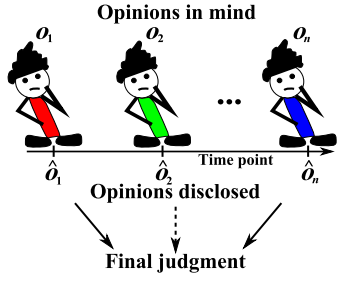


Figure 1: Collecting dependent opinions from crowd workers. True opinion of the i^{th} worker might be O_i but it is disclosed as \hat{O}_i (due to bias) after seeing the others' opinions $\{\hat{O}_1, \hat{O}_2, \dots, \hat{O}_{i-1}\}$.

crowd workers. Thus, it yields the opinions (annotation scores) of crowd reviewers on the said research materials. Using an online platform, the reviews are basically collected (Bhattacharyya 2014). For this specific platform, the possible options for opinions are 'strong accept', 'accept', 'weak accept', 'borderline', 'weak reject', 'reject', and 'strong reject'. However, such models can be used for any open crowdsourcing platform where the crowd workers provide a dependent nature of feedback.

Problem Formulation

Let us now formalize the dependent judgment analysis problem for a crowd-powered environment. Suppose we have a set of m questions $Q = \{q_1, q_2, \dots, q_m\}$ that are labeled by n number of crowd workers $W = \{w_1, w_2, \dots, w_n\}$. Further assume that the set of optional labels to be given as opinions is given by $L = \{l_1, l_2, \dots, l_k\}$. Given these precursory assumptions, an annotation process is a quadruplet (Q, W, τ, T) consisting of a mapping function $\tau : Q \times W \rightarrow L \times T$ and a set of arrival times. The objective is to obtain the final judgment of all the questions in Q . Fig. 1 depicts one such scenario where the dependent opinions arrive successively over time. Note that, $\hat{O}_1 = O_1$. Note that, the order of arrival of crowd workers is the only important factor here, not the distribution of arrival.

Challenges

It can be probabilistically understood that revealing the true opinion of a crowd worker from the given set of dependent opinions is not an easy task. Following the conditional probability, we have $P(O_i = \hat{o}_i | O_1 = \hat{o}_1, \dots, O_{i-1} = \hat{o}_{i-1})$

$$= \frac{P(O_1 = \hat{o}_1, \dots, O_{i-1} = \hat{o}_{i-1} | O_i = \hat{o}_i)}{P(O_1 = \hat{o}_1, \dots, O_{i-1} = \hat{o}_{i-1})}. \quad (1)$$

So, it might be useful to take the opinions in two different phases from the crowd workers to understand the dependence of the opinions from two different orders. Initially, the crowd workers provide their opinions independently (prior independent score) and these independent opinions are not disclosed to the others. In the next phase, all these independent opinions are made public so that every crowd worker

can observe others' independent scores. Then again, the opinions are obtained (posterior dependent score) from them on the same questions. This might provide a better understanding about the distribution of the actual opinions.

We now introduce a few terms that might be useful for a better implementation of the said ideas in the form of a judgment analysis model.

Drop of Confidence: The *drop of confidence* of a crowd worker is the absolute difference between the prior independent score and posterior dependent score. If the mean score changes (i.e., majority of the crowd workers deviate from their initial opinions), any arbitrary crowd worker is likely to change his posterior dependent score. So, how a crowd worker is modifying his own score with respect to the others for a particular question can be expressed as the ratio of deviation of individual score with respect to the deviation of mean score in independent and dependent situations, respectively. The reliability of a crowd worker can be quantified as the reciprocal of this ratio.

Reliability: The *reliability* of a crowd worker can be defined in terms of the drop of confidence of a crowd worker in comparison with the mean deviation of two different situations (i.e., independent and dependent). It reflects how much confident the crowd worker is over his own opinions. When a crowd worker changes his score drastically after viewing others' scores, he might not be confident enough. Thus, some bias gets included into his dependent opinion. On the other side, if a crowd worker sticks to his own score even after seeing the independent scores of others, the crowd worker is confident about his opinion. However, only becoming confident, while giving an opinion, does not guarantee that the worker is good. The absolute difference between his posterior dependent score and the mean value of all the posterior dependent scores should also be less. This gap highlights the closeness of a crowd worker with others on the basis of their posterior dependent scores.

Accuracy: The *accuracy* of a crowd worker is the closeness of his posterior dependent score to the mean of all posterior dependent scores over a particular question. This means how much similar a crowd worker's opinion is with rest of the crowd workers. A better accuracy score indicates that the crowd worker is highly similar to the other crowd workers. So, increasing this accuracy score with a less drop of confidence means a crowd worker is reliable and accurate.

Conclusions

We introduce a new judgment analysis problem that considers both independent and dependent opinions of crowd workers. We also discuss the requirements for developing an appropriate consensus model for managing dependent opinions of crowd workers along with independent crowd opinions. Consideration of the ambiguity of questions (for which confidence may drop) and the self-reported confidence scores of crowd workers (collected with annotations) can make the model more robust. It is also interesting to consider that a subset of existing opinions are only available, instead of all the opinions. Consideration of a paid crowdsourcing model is also encouraging.

Acknowledgments

The work of Malay Bhattacharyya is supported by the Visvesvaraya Young Faculty Research Fellowship 2015-16 of DeitY, Government of India. All the authors would like to thank the crowd contributors involved in this work.

References

- [Bhattacharyya 2013] Bhattacharyya, M. 2013. Opinion ensembling: Learning from dependent judgements of the crowd. In *Proceedings of the Crowdscale Shared Task Challenge*.
- [Bhattacharyya 2014] Bhattacharyya, M. 2014. Viability of crowd-volunteered open research reviews. In *Proceedings of the AAAI conference of Human Computation and Crowdsourcing (HCOMP) workshop on Citizen + X*, 6–7. Pittsburgh, USA: AAAI Press.
- [Chatterjee and Bhattacharyya 2015] Chatterjee, S., and Bhattacharyya, M. 2015. A biclustering approach for crowd judgment analysis. In *Proceedings of the 2nd ACM IKDD Conference on Data Sciences*, 118–119. Bangalore, India: ACM Press.
- [Jung and Lease 2012] Jung, H. J., and Lease, M. 2012. Improving quality of crowdsourced labels via probabilistic matrix factorization. In *Proceedings of the 4th Human Computation Workshop at AAAI*, 101–106. Toronto, Ontario, Canada: AAAI Press.
- [Kajino, Tsuboi, and Kashima 2012] Kajino, H.; Tsuboi, Y.; and Kashima, H. 2012. A convex formulation for learning from crowds. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 73–79. Toronto, Ontario, Canada: AAAI Press.
- [Kamar, Kapoor, and Horvitz 2015] Kamar, E.; Kapoor, A.; and Horvitz, E. 2015. Identifying and accounting for task-dependent bias in crowdsourcing. In *Third AAAI Conference on Human Computation and Crowdsourcing*, 92–101. San Diego, USA: AAAI Press.
- [Krishnan et al. 2014] Krishnan, S.; Patel, J.; Franklin, M. J.; and Goldberg, K. 2014. A methodology for learning, analyzing, and mitigating social influence bias in recommender systems. In *Proceedings of the 8th ACM Conference on Recommender Systems*, 137–144. Silicon Valley, USA: ACM.
- [Prelec 2004] Prelec, D. 2004. A bayesian truth serum for subjective data. *Science* 306(5695):462–466.
- [Raykar and Yu 2011] Raykar, V. C., and Yu, S. 2011. Eliminating spammers and ranking annotators for crowdsourced labeling tasks. *Journal of Machine Learning Research* 13:491–518.
- [Sheshadri and Lease 2013] Sheshadri, A., and Lease, M. 2013. SQUARE: A benchmark for research on computing crowd consensus. In *Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing*, 2035–2043. Palm Springs, USA: AAAI Press.